

WO 03/073839

1/9

PCT/IT03/00120

SEQUENCE LISTING

SEQ ID NO 1:

cDNA nucleotide sequence of the GCB gene.

Underlined sequence: signal peptide 1-57

Mature peptide: 58-1548

```

1      atggc tggcagcctc acaggtttgc ttctacttca ggcagtgtcg tgggcatcag
56     gtgcccgcct ctgcatccct aaaagcttcg gctacagctc ggtggtgtgt
106    gtctgcaatg ccacatactg tgactccttt gacccccga cctttcctgc
156    ccttggtacc ttcagccgct atgagagtag acgcagtggg cgacggatgg
206    agctgagtat ggggcccata caggctaata acacgggcac aggcctgcta
256    ctgaccctgc agccagaaca gaagttccag aaagtgaagg gatttggagg
306    ggccatgaca gatgctgctg ctctcaacat ccttgccctg tcacccctgc
356    cccaaaattt gctacttaaa tcgtacttct ctgaagaagg aatcgatat
406    aacatcatcc gggtagccat ggccagctgt gacttctcca tccgcaccta
456    cacctatgca gacaccctgc atgatttcca gttgcacaac ttcagcctcc
506    cagaggaaga taccaagctc aagatacccc tgattcacgc agccctgcag
556    ttggcccagc gtcccgtttc actccttgcc agcccctgga catcaccac
606    ttggctcaag accaatggag cgggtgaatgg gaaggggtca ctcaaggagc
656    agcccggaga catctaccac cagacctggg ccagatactt tgtgaagtgc
706    ctggatgcct atgctgagca caagttacag ttctgggcag tgacagctga
756    aaatgagcct tctgctgggc tgttgagtgg atacccttc cagtgcctgg
806    gcttcacccc tgaacatcag cgagacttca ttgccctga cctaggtcct
856    accctcgcca acagtactca ccacaatgtc cgcctactca tgctggatga
906    ccaacgcttg ctgctgcccc actgggcaaa ggtggtactg acagaccag
956    aagcagctaa atatgttcat ggcattgctg tacattggtc cctggacttt
1006   ctggctccag ccaaagccac cctaggggag acacaccgcc tgttcccaa
1056   caccatgctc tttgcctcag aggcctgtgt gggctccaag ttctgggagc
1106   agagtgtgcg gctaggctcc tgggatcgag ggatgcagta cagccacagc
1156   atcatcacga acctcctgta ccatgtggtc ggctggaccg actggaacct
1206   tgccctgaac ccgaaggag gacccaattg ggtgcgtaac tttgtcgaca
1256   gtcccatcat tgtagacatc accaaggaca cgttttataa acagcccatg
1306   ttctaccacc ttggccactt cagcaagtgc attcctgagg gctccagag
1356   agtggggctg gttgccagtc agaagaacga cctggacgca gtggcactga
1406   tgcacccga tggctctgct gttgtggtcg tgctaaaccg ctctctaa
1456   gatgtgcctc ttaccatcaa ggatcctgct gtgggcttcc tggagacaat
1506   ctacacctgg tactccattc acacctacct gtggcatcgc cagtga

```

SEQ ID NO 2: Amino acid sequence codifying human GCB and native signal peptide.

Met Ala Gly Ser Leu Thr Gly Leu Leu Leu Leu Gln Ala Val Ser Trp
Ala Ser Gly Ala Arg Pro Cys Ile Pro Lys Ser Phe Gly Tyr Ser Ser
Val Val Cys Val Cys Asn Ala Thr Tyr Cys Asp Ser Phe Asp Pro Pro
Thr Phe Pro Ala Leu Gly Thr Phe Ser Arg Tyr Glu Ser Thr Arg Ser
Gly Arg Arg Met Glu Leu Ser Met Gly Pro Ile Gln Ala Asn His Thr
Gly Thr Gly Leu Leu Leu Thr Leu Gln Pro Glu Gln Lys Phe Gln Lys
Val Lys Gly Phe Gly Gly Ala Met Thr Asp Ala Ala Ala Leu Asn Ile
Leu Ala Leu Ser Pro Pro Ala Gln Asn Leu Leu Leu Lys Ser Tyr Phe
Ser Glu Glu Gly Ile Gly Tyr Asn Ile Ile Arg Val Pro Met Ala Ser
Cys Asp Phe Ser Ile Arg Thr Tyr Thr Tyr Ala Asp Thr Pro Asp Asp
Phe Gln Leu His Asn Phe Ser Leu Pro Glu Glu Asp Thr Lys Leu Lys
Ile Pro Leu Ile His Arg Ala Leu Gln Leu Ala Gln Arg Pro Val Ser
Leu Leu Ala Ser Pro Trp Thr Ser Pro Thr Trp Leu Lys Thr Asn Gly
Ala Val Asn Gly Lys Gly Ser Leu Lys Gly Gln Pro Gly Asp Ile Tyr
His Gln Thr Trp Ala Arg Tyr Phe Val Lys Phe Leu Asp Ala Tyr Ala
Glu His Lys Leu Gln Phe Trp Ala Val Thr Ala Glu Asn Glu Pro Ser
Ala Gly Leu Leu Ser Gly Tyr Pro Phe Gln Cys Leu Gly Phe Thr Pro
Glu His Gln Arg Asp Phe Ile Ala Arg Asp Leu Gly Pro Thr Leu Ala
Asn Ser Thr His His Asn Val Arg Leu Leu Met Leu Asp Asp Gln Arg
Leu Leu Leu Pro His Trp Ala Lys Val Val Leu Thr Asp Pro Glu Ala
Ala Lys Tyr Val His Gly Ile Ala Val His Trp Tyr Leu Asp Phe Leu
Ala Pro Ala Lys Ala Thr Leu Gly Glu Thr His Arg Leu Phe Pro Asn
Thr Met Leu Phe Ala Ser Glu Ala Cys Val Gly Ser Lys Phe Trp Glu
Gln Ser Val Arg Leu Gly Ser Trp Asp Arg Gly Met Gln Tyr Ser His
Ser Ile Ile Thr Asn Leu Leu Tyr His Val Val Gly Trp Thr Asp Trp
Asn Leu Ala Leu Asn Pro Glu Gly Gly Pro Asn Trp Val Arg Asn Phe
Val Asp Ser Pro Ile Ile Val Asp Ile Thr Lys Asp Thr Phe Tyr Lys
Gln Pro Met Phe Tyr His Leu Gly His Phe Ser Lys Phe Ile Pro Glu
Gly Ser Gln Arg Val Gly Leu Val Ala Ser Gln Lys Asn Asp Leu Asp
Ala Val Ala Leu Met His Pro Asp Gly Ser Ala Val Val Val Val Leu
Asn Arg Ser Ser Lys Asp Val Pro Leu Thr Ile Lys Asp Pro Ala Val
Gly Phe Leu Glu Thr Ile Ser Pro Gly Tyr Ser Ile His Thr Tyr Leu
Trp His Arg Gln

SEQ ID NO 3: forward primer for GCB amplification

5': tctagaatggctggcagcctcacaggt

SEQ ID NO 4: reverse primer for GCB amplification

5': gtgtggatggacaccgtagcggtcactctcgag

SEQ ID NO 5: forward primer for GCB amplification

5': cccgggtgcccgcccctgcatccctaaaagc

SEQ ID NO 6: PGLOB promoter

```
1  taaaataatc tatacattaa aaaatttgat tttaaaattt tagaaattca tgattttatt
61  tttttttacc agaaatccgt taatattggt aaaatattac caactaattt ataaatttta
121 ttttaaggca attaagcatg tttgataaaa tatatatatt gttataaata cttttcaaaa
181 gtataaagtt gatgatggcg tgggtgtaga ttattttagt tctaggttcg aatgcaagtt
241 ggtttagaca tttagcctta ttcttttttc taaccaaagt aaatgtaaat ggaaaacctt
301 taggaaaaaa aagaaatcaa aattgaaaac atcatccggt ggagtcgaga agcccacacc
361 cacgtgaccc aacaatatta aaataagagt ttgctctaca gtaaatgcga tactttttta
421 ttcaatactt tttccacttc taaaatcttg gagatttgca ccgttaacta attaagtgtt
481 atatccaacg gtcctaaaaa aacttggtga ccgtgcctca catttcaact ttgcgcaccc
541 tagaagccgt ctatgttttag gttagtgttt gcaacagttg aagcgcatca ctcaggaggc
601 tacttgggtc tgcttttgcg tcttttggtc aatttttcac gtgattttgt tgggtgaacac
661 gcgtacttga aacttattat aaattacata attttataag tttcacttct tatataatac
721 ttcattcatg catttataat tttgatgaat aataaagagt ttgttaaaaa atatattatt
781 tcatataata tatagggttt agaatgccaa tttttaaaaa aagaataaaa aaataaatag
841 aataaaatcg aaaaaatgaa atgtaaaaaa tttgaggggg acaataaaaa tatgaaagtc
901 tattatttaa attttccatt agaattctat tttccttagt taatatgagc tagccagttg
961 ggagatacac gaaaatgtca tgaaacagtt gcatgtaggg aaattaatgt agtagagggg
1021 tagcaagaca aaaatccaag ccaagctagc tgctcacgag aactcgatcc acacgtcctt
1081 tacagagttt caaacggatg aaatctgcat ggcattgcaac taaagcattg ttctcagctg
1141 ccaagtaccc ctcacactca ccaacccttt gtttttctcc ccattgcatg ttaactcaag
1201 tttatccttt ctttgcttct ggaaatttca caagcctcaa acacgtcgac gtccaatctt
1261 gtgaccaaca cggccaaaag aaaagagaat ctcacccggt tcacacttag ccacttaag
1321 ctagccaaac ggtgatcttt ctctatatat tgtagctctc taacacaacc aacactacca
1381 ttattcaata ttcaaaccct gctctatact acacacacta gaagaata
```

SEQ ID NO 7: Soy basic glubulin 7S signal sequence

1 atggttctat cctccactac tttttagccc tctctctttc ttgctctttt cttttcttct
61 tatccgactc a

SEQ ID NO 8: cDNA nucleotide sequence of the GLA gene

Underlined sequence: signal peptide 21-116

Sequece coding for the mature peptide 117-1310

1 aatgctgtcc ggtcaccgtg acaatgcagc tgaggaaccc agaactacat ctgggctgcg
61 cgcttgcgct tcgcttcctg gccctcgttt cctgggacat ccctggggct agagcactgg
121 acaatggatt ggcaaggacg cctaccatgg gctggctgca ctgggagcgc ttcattgtga
181 accttgactg ccaggaagag ccagattcct gcatcagtga gaagctcttc atggagatgg
241 cagagctcat ggtctcagaa ggctggaagg atgcaggtta tgagtacctc tgcattgatg
301 actgttggat ggctccocaa agagattcag aaggcagact tcaggcagac cctcagcgct
361 ttcctcatgg gattcgccag ctagctaatt atgttcacag caaaggactg aagctaggga
421 tttatgcaga tgttgaaaat aaaacctgcg caggcttccc tgggagtttt ggatactacg
481 acattgatgc ccagaccttt gctgactggg gtagtagatct gctaaaattt gatggttgtt
541 actgtgacag tttgaaaaat ttggcagatg gttataagca catgtccttg gccctgaata
601 ggactggcag aagcattgtg tactcctgtg agtggcctct ttatatgtgg ccctttcaaa
661 agcccaatta tacagaaatc cgacagtact gcaatcactg gcgaaatttt gctgacattg
721 atgattcctg gaaaagtata aagagtatct tggactggac atcttttaac caggagagaa
781 ttgttgatgt tgctggacca gggggttggg atgacccaga tatgttagtg attggcaact
841 ttggcctcag ctggaatcag caagtaactc agatggccct ctgggctatc atggctgctc
901 ctttattcat gtctaattgac ctccgacaca tcagccctca agccaaagct ctcttcagg
961 ataaggacgt aattgccatc aatcaggacc ccttgggcaa gcaagggtac cagcttagac
1021 agggagacaa ctttgaagtg tgggaacgac ctctctcagg cttagcctgg gctgtagcta
1081 tgataaacccg gcaggagatt ggtggacctc gctcttatac catcgagtt gcttccctgg
1141 gtaaaggagt ggctgtaat cctgcctgct tcatcacaca gctcctccct gtgaaaagga
1201 agctagggtt ctatgaatgg acttcaaggt taagaagtca cataaatccc acaggcactg
1261 ttttgcttca gctagaaaat acaatgcaga tgtcattaaa agacttactt taaaatgtt

SEQ ID NO 9: Amino acid sequence coding for GLA and native
signal peptide

Thr Met Gln Leu Arg Asn Pro Glu Leu His Leu Gly Cys Ala Leu Ala
Leu Arg Phe Leu Ala Leu Val Ser Trp Asp Ile Pro Gly Ala Arg Ala
Leu Asp Asn Gly Leu Ala Arg Thr Pro Thr Met Gly Trp Leu His Trp
Glu Arg Phe Met Cys Asn Leu Asp Cys Gln Glu Glu Pro Asp Ser Cys

Ile Ser Glu Lys Leu Phe Met Glu Met Ala Glu Leu Met Val Ser Glu
 Gly Trp Lys Asp Ala Gly Tyr Glu Tyr Leu Cys Ile Asp Asp Cys Trp
 Met Ala Pro Gln Arg Asp Ser Glu Gly Arg Leu Gln Ala Asp Pro Gln
 Arg Phe Pro His Gly Ile Arg Gln Leu Ala Asn Tyr Val His Ser Lys
 Gly Leu Lys Leu Gly Ile Tyr Ala Asp Val Gly Asn Lys Thr Cys Ala
 Gly Phe Pro Gly Ser Phe Gly Tyr Tyr Asp Ile Asp Ala Gln Thr Phe
 Ala Asp Trp Gly Val Asp Leu Leu Lys Phe Asp Gly Cys Tyr Cys Asp
 Ser Leu Glu Asn Leu Ala Asp Gly Tyr Lys His Met Ser Leu Ala Leu
 Asn Arg Thr Gly Arg Ser Ile Val Tyr Ser Cys Glu Trp Pro Leu Tyr
 Met Trp Pro Phe Gln Lys Pro Asn Tyr Thr Glu Ile Arg Gln Tyr Cys
 Asn His Trp Arg Asn Phe Ala Asp Ile Asp Asp Ser Trp Lys Ser Ile
 Lys Ser Ile Leu Asp Trp Thr Ser Phe Asn Gln Glu Arg Ile Val Asp
 Val Ala Gly Pro Gly Gly Trp Asn Asp Pro Asp Met Leu Val Ile Gly
 Asn Phe Gly Leu Ser Trp Asn Gln Gln Val Thr Gln Met Ala Leu Trp
 Ala Ile Met Ala Ala Pro Leu Phe Met Ser Asn Asp Leu Arg His Ile
 Ser Pro Gln Ala Lys Ala Leu Leu Gln Asp Lys Asp Val Ile Ala Ile
 Asn Gln Asp Pro Leu Gly Lys Gln Gly Tyr Gln Leu Arg Gln Gly Asp
 Asn Phe Glu Val Trp Glu Arg Pro Leu Ser Gly Leu Ala Trp Ala Val
 Ala Met Ile Asn Arg Gln Glu Ile Gly Gly Pro Arg Ser Tyr Thr Ile
 Ala Val Ala Ser Leu Gly Lys Gly Val Ala Cys Asn Pro Ala Cys Phe
 Ile Thr Gln Leu Leu Pro Val Lys Arg Lys Leu Gly Phe Tyr Glu Trp
 Thr Ser Arg Leu Arg Ser His Ile Asn Pro Thr Gly Thr Val Leu Leu
 Gln Leu Glu Asn Thr Met Gln Met Ser Leu Lys Asp Leu Leu

SEQ ID NO 10: Forward primer for GLA amplification

5': ggatccctggacaatggattggcaaggac

SEQ ID NO 11: Reverse primer for GLA amplification

5': gtctacagtaattttctgaatgaaattctatag

SEQ ID NO 12: cDNA GAA.

Underlined sequence: signal peptide 220-426

Sequence coding for the mature peptide 427-3075

1 cagttgggaa agctgaggtt gtcgccgggg ccgcgggtgg aggtcgggga tgaggcagca
 61 ggtaggacag tgacctcggg gacgcgaagg accccggcca cctctaggtt ctctcgtcc

121 gcccgttgtt cagcgaggga ggctctgggc ctgccgcagc tgacggggaa actgaggcac
181 ggagcgggcc tgtaggagct gtccaggcca tctccaacca tgggagtga gcacccgccc
241 tgctcccacc ggctcctggc cgtctgcgcc ctcgtgtcct tggcaaccgc tgactcctg
301 gggcacatcc tactccatga tttcctgctg gttccccgag agctgagtgg ctcctcccca
361 gtcctggagg agactcacc agctcaccag cagggagcca gcagaccagg gccccgggat
421 gcccaggcac accccggccg tcccagagca gtgcccacac agtgcgacgt ccccccaac
481 agccgcttcg attgcgcccc tgacaaggcc atcaccagg aacagtgcga gggccgggc
541 tgctgctaca tccctgcaaa gcaggggctg cagggagccc agatggggca gccctggtgc
601 ttcttccac ccagctaccc cagctacaag ctggagaacc tgagctctc tgaatgggc
661 tacacggcca ccctgacccg taccaccccc accttcttcc ccaaggacat cctgaccctg
721 cggctggacg tgatgatgga gactgagaac cgctccact tcacgatcaa agatccagct
781 aacaggcgct acgaggtgcc cttggagacc ccgctgtcc acagccgggc accgtcccca
841 ctctacagcg tggagtctc cgaggagccc ttcgggtga tcgtgcaccg gcagctggac
901 ggccgcgtgc tgctgaacac gacggtggcg cccctgttct ttgcggacca gttccttcag
961 ctgtccacct cgctgccctc gcagtatatc acaggcctcg ccgagcacct cagtccctg
1021 atgctcagca ccagctggac caggatcacc ctgtggaacc gggaccttgc gccacgcgc
1081 ggtgcgaacc tctacgggtc tcacctttc tacctggcgc tggaggacgg cgggtcggca
1141 cacggggtgt tctgctaaa cagcaatgcc atggatgtgg tctgcagcc gagccctgcc
1201 cttagctgga ggtcgacagg tgggatcctg gatgtctaca tcttctggg cccagagccc
1261 aagagcgtgg tgcagcagta cctggacgtt gtgggatacc cgttcatgcc gccatactgg
1321 ggcttgggct tccacctgtg ccgctggggc tactcctcca ccgctatcac ccgccagtg
1381 gtggagaaca tgaccagggc ccacttcccc ctggacgtcc aatggaacga cctggactac
1441 atggactccc ggagggactt cacgttcaac aaggatggct tccgggactt cccggccatg
1501 gtgcaggagc tgcaccaggg cggccggcgc tacatgatga tcgtggatcc tgccatcagc
1561 agctcgggcc ctgccgggag ctacaggccc tacgacgagg gtctgcggag gggggttttc
1621 atcaccaacg agaccggcca gccgctgatt gggaaggat gggccgggtc cactgccttc
1681 cccgacttca ccaacccccc agccctggcc tgggtgggagg acatggtggc tgagttccat
1741 gaccagggtgc cttcgcagg catgtggatt gacatgaacg agccttccaa cttcatcaga
1801 ggctctgagg acggctgccc caacaatgag ctggagaacc caccctacgt gcctggggtg
1861 gttgggggga ccctccagge ggccaccatc tgtgcctcca gccaccagtt tctctccaca
1921 cactacaacc tgcacaacct ctacggcctg accgaagcca tcgcctcca cagggcgctg
1981 gtgaaggctc gggggacacg ccatttgtg atctcccgt cgaccttgc tggccacggc
2041 cgatacgccg gccactggac gggggacgtg tggagctcct gggagcagct cgcctcctcc
2101 gtgccagaaa tctgcagtt taacctgctg ggggtgcctc tggtcggggc cgacgtctgc
2161 ggcttctctg gcaacacctc agaggagctg tgtgtgcgct ggaccagct gggggccttc
2221 tacccttca tgcggaacca caacgcctg ctcatgtctc cccaggagcc gtacagcttc
2281 agcgagccgg cccagcagge catgaggaag gccctcacc tgctgtacgc actcctcccc

2341 cacctctaca cactgttcca ccaggcccac gtgcgggggg agaccgtggc ccggcccctc
2401 ttcctggagt tccccaagga ctctagcacc tggactgtgg accaccagct cctgtggggg
2461 gaggccctgc tcatcacccc agtgctccag gccgggaagg ccgaagtga ctggtacttc
2521 cccttgggca catggtacga cctgcagacg gtgccaatag aggcccttgg cagcctccca
2581 ccccccacctg cagctccccg tgagccagcc atccacagcg aggggcagtg ggtgacgctg
2641 ccggccccc tggacaccat caacgtccac ctccgggctg ggtacatcat cccctgacg
2701 ggccctggcc tcacaaccac agagtccgc cagcagccca tggccctggc tgtggccctg
2761 accaagggtg gagaggcccg aggggagctg ttctgggacg atggagagag cctggaagtg
2821 ctggagcgag gggcctacac acaggtcatc ttctggcca ggaataacac gatcgtgaat
2881 gagctgttac gtgtgaccag tgagggagct ggcctgcagc tgcagaaggt gactgtcctg
2941 ggcgtggcca cggcgcccc gacaggtcctc tccaacggtg tccctgtctc caacttcacc
3001 tacagccccg acaccaaggt cctggacatc tgtgtctcgc tgttgatggg agagcagttt
3061 ctcgctcagct ggtgttagcc gggcgaggagtg tgttagtctc tccagaggga ggctggttcc
3121 ccagggaagc agagcctgtg tgcgggcagc agctgtgtgc gggcctgggg gttgcatgtg
3181 tcacctggag ctgggacta accattccaa gccgccgcat cgcttgtttc cacctcctgg
3241 gccggggctc tggcccccaa cgtgtctagg agagctttct ccctagatcg cactgtgggc
3301 cggggcctgg agggctgctc tgtgttaata agattgtaag gtttgccctc ctacactgtt
3361 gccggcatgc gggtagtatt agccaccccc ctccatctgt tcccagcacc ggagaagggg
3421 gtgctcaggt ggaggtgtgg ggtatgcacc tgagctcctg cttcgcgctt gctgctctgc
3481 cccaacgcga ccgcttcccg gctgccaga gggctggatg cctgccggtc cccgagcaag
3541 cctgggaact caggaaaatt cacaggactt gggagattct aaatcttaag tgcaattatt
3601 ttaataaaaag gggcatttgg aatc

SEQ ID NO 13: Amino acid sequence of human GAA and native
signal peptide

Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys Asp Val Pro
Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile Thr Gln Glu
Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys Gln Gly Leu
Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr
Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr
Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu
Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg Leu His Phe
Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr
Pro Arg Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe
Ser Glu Glu Pro Phe Gly Val Ile Val His Arg Gln Leu Asp Gly Arg
Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe
Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala

Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr
Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly
Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly
Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser
Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile
Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val
Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu
Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu
Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu
Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe
Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly Gly Arg Arg
Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly
Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr
Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr
Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp
Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly Met Trp Ile
Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys
Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly
Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu
Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile
Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val
Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp
Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro
Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp
Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp
Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu
Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln
Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu
Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr Val Ala Arg
Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp
His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln
Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr
Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro
Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val
Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly
Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg
Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala

Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu
Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile
Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu
Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu
Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys
Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val
Ser Trp Cys

SEQ ID NO 14: Forward primer for GAA amplification

5': gatatctgcacaccccgccgtcccag

SEQ ID NO 15: Reverse primer for GAA amplification

5': gtcaaagagcagtcgaccacaatcctatag